Evaluability Assessment to Improve Public Health Policies, Programs, and Practices*

Laura C. Leviton,[1] Laura Kettel Khan,[2] Debra Rog,[3] Nicola Dawkins,[4] and David Cotton[4]

[1] The Robert Wood Johnson Foundation, Princeton, New Jersey 08543-2316; email: llevito@rwjf.org

[2] Division of Nutrition, Physical Activity and Obesity, Centers for Disease Control and Prevention, Atlanta, Georgia 30341; email: ldk7@cdc.gov

[3] The Rockville Institute and Westat, Incorporated, Rockville, Maryland 20850; email: DebraRog@RockvilleInstitute.org

[4] ICF Macro, Atlanta, Georgia 30329; email: DCotton@orcmacro.com, nicola.u.dawkins@orcmacro.com

*The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Key Words

exploratory evaluation, logic model, performance measurement, public health programs, stakeholders, theory of change

Abstract

Evaluability assessment, also commonly known as exploratory evaluation, has assisted the field of public health to improve programs and to develop a pragmatic, practice-based research agenda. Evaluability assessment was originally developed as a low-cost pre-evaluation activity to prepare better for conventional evaluations of programs, practices, and some policies. For public health programs, however, it serves several other important purposes: (*a*) giving program staff rapid, constructive feedback about program operations; (*b*) assisting the core public health planning and assurance functions by helping to develop realistic objectives and providing low-cost, rapid feedback on implementation; (*c*) navigating federal performance measurement requirements; (*d*) translating research into practice by examining the feasibility, acceptability, and adaptation of evidence-based practices in new settings and populations; and (*e*) translating practice into research by identifying promising new approaches to achieve public health goals.

OVERVIEW

Definition

Evaluability assessment is a pre-evaluation activity designed to maximize the chances that any subsequent evaluation of programs, practices, or policies will result in useful information (47). Throughout this review, we refer to public health programs, but the method is equally relevant to public health practices, program components, and many policies. First employed in the mid-1970s by the Department of Health, Education and Welfare (HEW), the forerunner of the Department of Health and Human Services (HHS), evaluability assessment is experiencing a resurgence of interest and practice (38, 43). However, we argue that the method should be used even more.

From its inception, evaluability assessment was also called exploratory evaluation. Although the term evaluability assessment is more commonly used today, exploratory evaluation is easier to say, and in our experience, program managers find it less threatening. Both terms should convey that this method is less than a complete evaluation; by itself it is not adequate as an evaluation. It does not have a sufficient standard of evidence to demonstrate effectiveness or describe implementation. However, it is valuable to identify programs that are unlikely to be effective in their current form, as well as those that show promise to address societal needs. It can assist program planners to adjust their activities and resources to achieve objectives, or to adjust their objectives in light of program reality.

In their review and critique of evaluation theories, Shadish et al. (41) define evaluability assessment as "assessing whether the program is ready to be managed for results, what changes are needed to do so, and whether the evaluation would contribute to improved program performance" (p. 225). Joseph Wholey (47), the originator of evaluability assessment, emphasizes the method as a better way to assure usefulness because it addresses "the objectives, expectations, and information needs of program managers and policy makers; explores program reality; assesses the likelihood that program activities will reach measurable progress toward program objectives; and assesses the extent to which evaluation information is likely to be used by program management" (p. xiii). It is one of very few systematic tools for planning evaluations. Although public health has developed other tools to plan evaluations, we demonstrate how they are complementary to, and do not replace, evaluability assessments.

Aim

This article describes how evaluability assessment has benefited public health and could do so in future. We describe the rationale, history, and evolution of evaluability assessment. We outline the steps in the method and distinguish it from related concepts. We then illustrate how evaluability assessment can benefit public health in five ways:

1. Serving the core public health functions of planning and assurance,
2. Building evaluation capacity,
3. Navigating federal performance measurement requirements,
4. Translating evidence-based research models into practice, and
5. Translating practice into research by identifying promising practices.

In spite of its merits, evaluability assessment is largely invisible in public health training, research, and practice. It is not even mentioned in some of the introductory evaluation textbooks that schools of public health commonly employ. It is our impression that academic public health tends to emphasize a fairly limited array of evaluation methods, including experimentation and quasi-experimentation on the one hand, and qualitative methods (often allied with community-based participatory research) on the other. Yet 40 years of experience indicates that effective evaluation makes flexible use of a variety of methods (41). Evaluability assessment is a useful precursor to both quantitative and qualitative evaluation.

Exploratory evaluation: evaluability assessment

Performance measurement: measurements, agreed to beforehand, that show progress on objectives or strategic plans

The Case for Cost-Effectiveness

Public health is usually strapped for resources, including evaluation resources, and cannot afford to waste them. Evaluability assessment offers a method that is likely to be cost-effective under many circumstances to improve public health programs, practices, and policy implementation. It is cost-effective because it sharpens the focus of programs and practices by eliciting the underlying logic model or theory of change, then examining whether resources and activities are sufficient and relevant to achieve significant change toward the desired objectives. If they are, then evaluation can proceed; if they are not, then either the resources and activities need to be adjusted or the objectives need to be revised. Of course, program evaluation itself can sometimes sharpen the focus of programs in these ways. However, evaluability assessment does so quickly, specifically, and at substantially lower cost than do most program evaluations.

Evaluability assessment is also cost-effective because it can prevent costly evaluation of programs and practices when the logic models are not plausible or when the programs still need to develop better and more relevant activities. Under these circumstances, evaluation is premature at best. Unfortunately, many millions of dollars have been spent on "no-effect" conclusions from the evaluation of health and social programs for precisely these reasons (7, 20, 35, 41, 47, 50). Finally, and perhaps most important, evaluability assessment can increase the relevance and usefulness of the evaluations that are conducted. As we demonstrate below, evaluability assessments can point the way to evaluation projects that have the best chance of revealing important information that people will use.

RATIONALE, HISTORY, AND EVOLUTION OF EVALUABILITY ASSESSMENT

Rationale: The Challenges and Opportunities of Program Reality

Evaluability assessment developed as a way to cope with several common challenges in program evaluation. (*a*) Evaluability assessment can shed light on disagreements among stakeholders, that is, those with an interest in the program or its evaluation. If disagreements persist, the program may not be ready for evaluation. (*b*) Often the logic underlying the program or practice has not been described. A good evaluability assessment can describe the program logic. If the program logic is not satisfactory, then the program is not plausible to achieve its result, and the staff need to make adjustments. (*c*) Staff may be able to describe program logic, but goals and objectives may not be realistic in light of resources provided or the activities being undertaken. When implementation is so deeply flawed, evaluation would be premature at best. Evaluability assessment can indicate the need for adjustments in activities and resources (a program development function) or the need for formative evaluation. (*d*) Stakeholders may agree about the goals and logic of a program, yet they may still not agree about performance criteria or how to measure program effectiveness. Evaluability assessment can inform stakeholders about options for evaluation and their potential usefulness. (*e*) The cost of an evaluation may be more than a program can afford, or (*f*) decision makers may be unwilling to make changes on the basis of evaluation. In both cases, evaluability assessment will reveal problems before decision makers commit to evaluation.

These challenges are pervasive and critically important. For example, many evaluations arrive at no-effect conclusions because the program in question was not fully developed, was fatally flawed, or was so puny in terms of the treatments actually being provided that a change in outcomes could never hope to be achieved (20, 35, 41, 50). Evaluations of outcome are also seriously impaired by a flawed design, measurement, or analysis—issues that could be addressed by better evaluation planning (20, 41, 50). To expend resources on such evaluations is wasteful and has been found to have a chilling effect on innovation (41). By avoiding such evaluations, evaluability assessment can help the field to adhere to

Logic model: a graphic depiction of the rationale and expectations of the program

Theory of change: describes the mechanisms through which the initiative's inputs and activities are thought to lead to the desired outcomes

Stakeholders: those with a stake or interest in the program or in its evaluation

Outcomes: achievement of short, intermediate, and long-term objectives of a program

the American Evaluation Association's *Guiding Principles for Evaluators* (1).

In conducting more than 100 evaluability assessments to date, we have noticed that they also help to protect the evaluator. Unfortunately people tend to blame the messenger, and this happens all too frequently to those who conduct evaluation (15). Evaluability assessment provides the evaluator with a paper trail to document what stakeholders said they wanted, and when they said it. It helps evaluators to formalize the agreements about the evaluation questions to be asked, the program being assessed, and the measures and design to be used. With such documentation, it becomes much more difficult for stakeholders to play revisionist games at the end of an evaluation with mixed findings or no-effect conclusions. All too frequently evaluators hear that "you did not understand our program" or "these outcome measures are irrelevant to what we are really trying to do." It is sad but true: Evaluability assessment helps keep the paper trails that are essential to preserve evaluators' reputations and even their jobs.

However, to focus only on the challenges is far too bleak an outlook—there are opportunities as well. On the positive side of the ledger is the use of evaluability assessment for program improvement. Investigators initially expected that evaluability assessment would result in better summative or outcome evaluations. However, evaluability assessment evolved as practitioners gained experience in using it. It serves as a valuable guide to program development (42). It also helps to clarify, and to inquire about, the assumptions underlying a program, an activity that is highly valued by clients (41). And far from being limited to outcome evaluations, it can guide the selection of a wide array of possible evaluation questions and methods (43).

A new opportunity for evaluability assessment to make a difference in public health is through its incorporation into the systematic screening and assessment (SSA) method, described below. The SSA method capitalizes on the rich stew of innovation that practitioners create as they are driven to solve public health problems. The SSA method does this by identifying the most promising innovations, whether they be programs, practices, or policies, then subjecting them to evaluability assessment. In the area of childhood obesity, many innovations in practice settings might never have come to evaluators' attention if the SSA method were not pursued as a first step in their evaluation (19).

History

Wholey and his colleagues at the Urban Institute developed evaluability assessment in the late 1970s in response to the challenge of dealing with programs that were unsuitable or premature for evaluation (29, 37, 47). In particular, federal programs were often evaluated against unrealistic goals; program design lacked logic; decision makers disagreed about the outcomes of importance and felt little ownership of the evaluation findings. Wholey instituted evaluability assessment most broadly at HEW in the 1970s and 1980s, and then at HHS. Wholey was the Deputy Assistant Secretary for Planning and Evaluation at HEW during the Carter administration. At HEW, the method was promoted as a way to expend the 1% of program funds that had been set aside for evaluation. Its use was probably inspired as much by the particular needs of the programs as by the desire to test the new tool. In reviewing the implementation of the method, Rog (37) found little variation in the study tasks or methods used but considerable variability in the costs and scope of the studies. Most of the studies provided options to improve program management, develop performance measures, and design evaluation strategies. Few were followed by subsequent evaluations, however.

The National Institute of Justice and the National Bureau of Standards also made substantial use of the method, and tools similar to evaluability assessment were used by the Food and Nutrition Service of the U.S. Department of Agriculture and by the Canadian

government (37, 40). In the late 1970s and 1980s, more than 50 evaluability assessments were conducted, 33 of these between 1980 and 1985. Wholey left government and use of the technique dropped off significantly. Between 1986 and 2000, only eight evaluability assessments could be identified (38).

Present-Day Uses of Evaluability Assessment

More recent tracking suggests that the method is reemerging (38, 43), owing in part to the increased federal emphasis on performance measurement. The passage of the Government Performance Results Act (GPRA) in 1993 signaled this renewed emphasis (31), and in 2002 the federal Office of Management and Budget (OMB) began requiring that agencies use the Program Assessment Rating Tool (PART) (30).

Recent reports can be seen in **Table 1**, which lists evaluability assessments of programs in public health, health care, mental health and substance abuse, and social services. More than 160 organizations at national, state, and local levels were involved in these assessments. The reports in **Table 1** are not an exhaustive listing, and most of them are summarized in review articles (11, 18, 19, 43). Many notable efforts are found in literature that is not peer-reviewed— books, reports, and meetings—because they are of primary concern to program managers and local stakeholders, not to the wider scientific community. For example, by 2004, the CDC had provided guidance on evaluability assessment in HIV prevention to 402 domestic and international government and nongovernmental organization staff (5). This effort was motivated by the need to increase both evaluation capacity and program development in HIV prevention. The CDC continues to provide this guidance through the HIV Prevention Program Evaluation Materials Database (28), and the perceived need for evaluability assessment influenced the way that community-based organizations were selected to implement a 13-city trial of HIV prevention (23).

DESCRIPTION OF EVALUABILITY ASSESSMENT

The Process

Wholey (48) outlined six steps for conducting evaluability assessments: involving the intended users of evaluation information, clarifying the intended program, exploring program reality, reaching agreement on needed changes in activities or goals, exploring alternative evaluation designs, and agreeing on evaluation priorities and intended uses of information. This description of steps appears overly linear in our experience. As seen in **Figure 1**, evaluability assessment is a cyclical, iterative process that builds understanding of the program design, the underlying program logic model or theory of change, opportunities for useful evaluation, and potential program improvement.

The initial step in evaluability assessment is to involve the potential evaluation users, obtaining their commitment to the project and defining the scope and purpose of the effort. The assessment team reviews program documents: vision and mission statements, descriptions, written goals and objectives if any, funded grant proposals, and others. Concurrently with this process, the assessment team interviews the stakeholders. Usually this process begins with one or more primary stakeholders, such as the program manager, policy maker responsible for program oversight, or the funder. Often an introductory meeting is useful to become oriented to the program, meet the primary stakeholders, and obtain documents. Other stakeholders include groups representing service participants or clients, representatives of the program staff or service providers, community groups and other institutions, and agencies that are affected by the program (34). For other stakeholder interviews, the evaluability assessment team will want to prepare by reviewing documents in detail. These early stages can be informative all by themselves because program stakeholders may not agree about what the program is supposed to accomplish. Disagreements will most often need to be resolved

Government Performance Results Act (GPRA): requires federal agencies measure their performance to demonstrate that they are using funds effectively

Office of Management and Budget (OMB): executive agency that advises the President on the federal budget

Program Assessment Rating Tool (PART): a method of performance measurement required by OMB

CDC: the U.S. Centers for Disease Control and Prevention (an agency of HHS)

Table 1 Recent evaluability assessments in public health and related areas

Discipline	Reference number(s)/source	N of sites
Public health programs in journals, books, reports, presentations		
AIDS outreach and prevention	18	1
Breastfeeding peer counselors	18	1
Childhood obesity–prevention policies and programs	19	48
Children's physical activity program	43	1
Community nursing for African American elders	18	1
Indian Health Service Injury Prevention Program	43	1
Mobile school health education program	18	1
Neighborhood violence prevention project	18	1
Prenatal program	48	1 state
Pilot childhood obesity–prevention programs	32, 36, 39	27
Rape prevention and education	43	1
Restaurant health promotion in Canada	43	1
Sexually transmitted disease control	18	1
Teen pregnancy prevention	18	1
Work site programs for obesity prevention	11	9
Public health programs: under way in 2009		
CDC cardiovascular disease–prevention projects	ICF Macro	At least 3
Robert Wood Johnson Foundation Games for Health Initiative	ICF Macro	1
Robert Wood Johnson Foundation Metaleadership in Public Health Initiative	ICF Macro	1
Trust for America's Health	Urban Institute	1
Health care programs in journals, books, reports, presentations		
Aquatic exercise therapy	18	1
Arthritis information service	18	1
Assessment of elderly at risk of falls	18	1
Canadian patient care and outcome process	43	1
Canadian training: complementary/alternative medicine	43	1
Geriatric education center	18	1
Herpes continuing medical education	18	1
Paramedic continuing education	18	1
Rapid response to needle stick	18	1
State organ donor program	18	1
Teen maternity clinic	18	1
Telemedicine	43	1
Health care system: under way in 2009		
Robert Wood Johnson Foundation Innovations in Nursing Education	Rutgers University	10
Mental health/substance abuse programs in journals, books, reports, presentations		
Canadian mental health services	43	National
Community Mental Health Block Grant Program	26	National
Harm reduction	43	1
Mental health programs	43	>1

(Continued)

Table 1 (*Continued*)

Discipline	Reference number(s)/source	N of sites
School psychology	43	1
Survivors of torture	43	1
Youth drug prevention	18	1
Social services in journals, books, reports, presentations		
Community-based programs	43	>1
Crime prevention	43	1
Discharge planning and the prevention of homelessness	24	8
Family preservation programs	48	National
Housing and community development partnerships	43	>1
Housing for people with AIDS	18	1
Infant care training program	18	1
Juvenile justice programs	13	National
Parenting education	43	1
Performance measurement and accountability	43	National
Refugee feeding centers	43	1
Social services: under way in 2009		
Robert Wood Johnson Foundation initiative on intimate partner violence in new immigrant populations	LTG Associates	10

before a full evaluation can be useful, but experience indicates that some disagreements (relating to self-interest or ideology) may never be resolved. Developing a logic model and scouting the reality of the program can still inform these debates.

On the basis of interviews and document review, the assessment team creates a logic model or theory of change and revises it continually as more information is gained. As it is developed, the logic model is shared with stakeholders for their reactions. It highlights areas of the program that are important to understand further as well as program assumptions that require verification. Once there is general agreement among stakeholders and evaluators about the logic model, evaluability assessment moves on to scouting the program reality as the assessment team observes activities and interviews a small sample of staff or clients. On this basis, the assessment team will again revise the logic model and share it. Program stakeholders finally receive a report that indicates the logic model's plausibility to achieve the desired goals and objectives, areas for further program improvement, the feasibility of conducting a full evaluation, options for evaluation design and questions that could be addressed, and a critique of the data that are currently being collected or might be collected.

Logic Models and Theories of Change

Central to the method is the creation of a logic model or theory of change. A logic model is a graphic depiction of the rationale and expectations of the program (22). A theory of change expands the logic model by describing the mechanisms through which the initiative's inputs and activities are thought to lead to the desired outcomes. The theory of change can rely on grand behavioral theories or prior empirical evidence, but it can also represent "small theory" about how the program is supposed to work (21).

Logic models and theories of change are often used outside of evaluability assessment to help in designing programs, guiding evaluations, and even guiding program management. In a recent survey, nonprofit

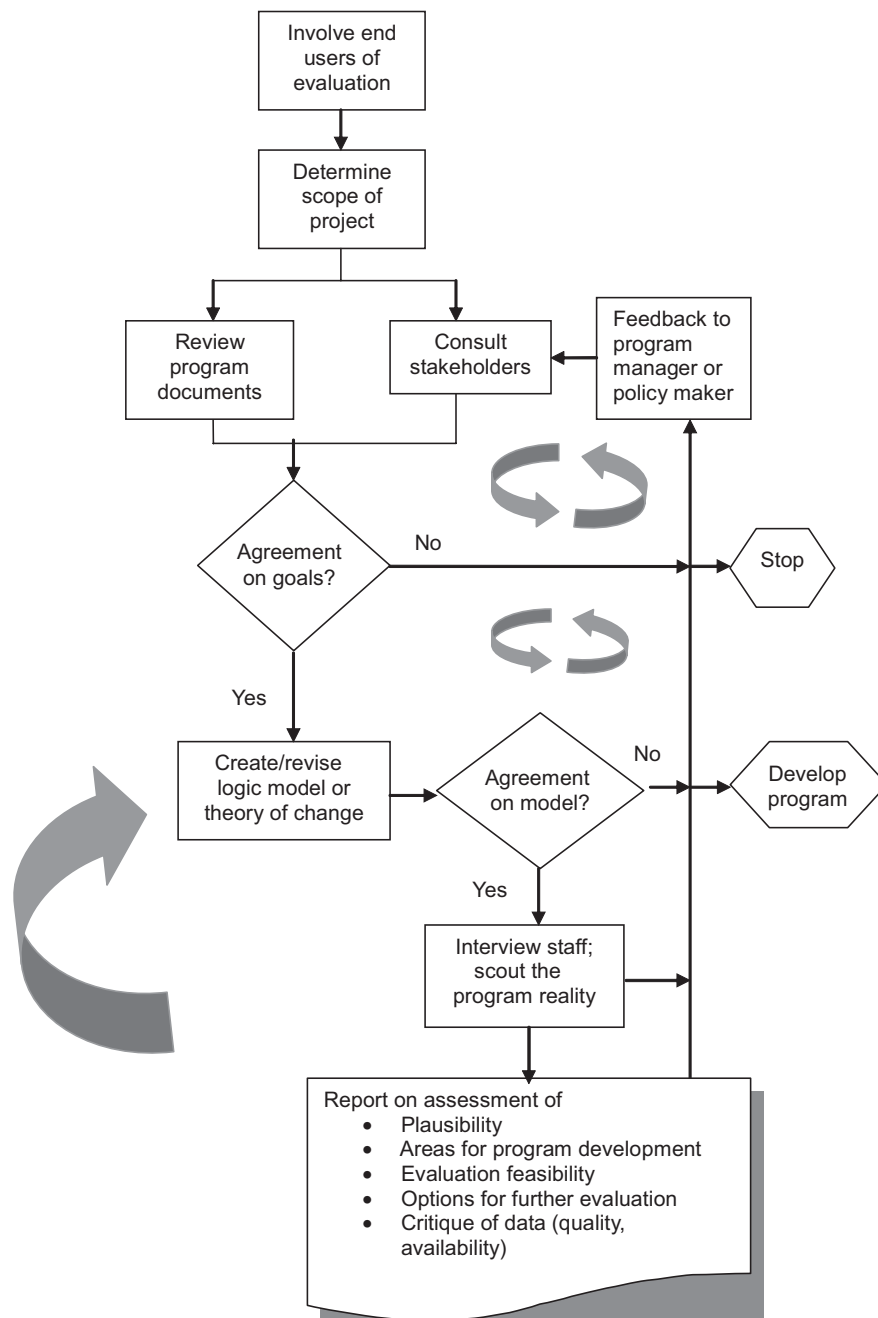**Inputs:** program resources, such as funding and personnel
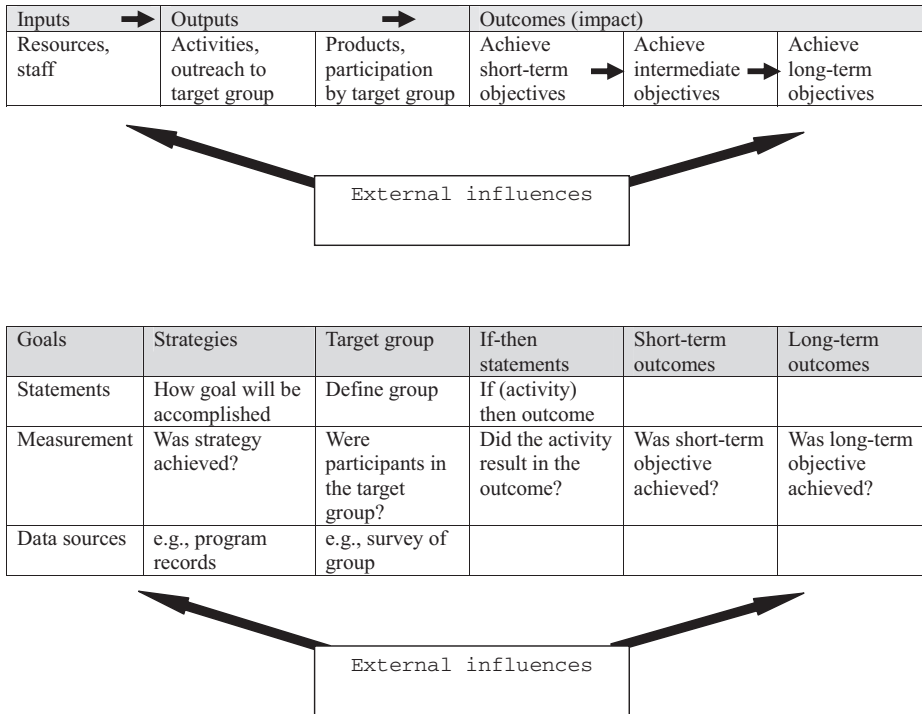
Figure 1

The evaluability assessment process.

Inputs	Outputs		Outcomes (impact)		
Resources, staff	Activities, outreach to target group	Products, participation by target group	Achieve short-term objectives	Achieve intermediate objectives	Achieve long-term objectives

External influences

Goals	Strategies	Target group	If-then statements	Short-term outcomes	Long-term outcomes
Statements	How goal will be accomplished	Define group	If (activity) then outcome		
Measurement	Was strategy achieved?	Were participants in the target group?	Did the activity result in the outcome?	Was short-term objective achieved?	Was long-term objective achieved?
Data sources	e.g., program records	e.g., survey of group			

External influences

Figure 2

Two generic logic models.

organizations named logic models as the most useful product of evaluation (2). In practice, we find few differences between graphic depictions that are called logic models and those termed theories of change. Logic models are not better than theories of change or vice versa. The best choice is the one that fits the needs of the stakeholders for planning useful evaluations.

The considerations for a best choice are illustrated in **Figure 2**. Two generic logic models are presented, but they are by no means the only choices: For example, flow charts are commonly used. Both of the generic models are read from left to right. Often the boxes will have arrows between them to convey a causal relationship; some arrows are omitted in this figure.

The first logic model begins with the inputs of the program such as resources (funding, equipment, space) and personnel (number, qualifications). Outputs are the direct result of the application of these resources by staff for activities (such as services), outreach to the target population, products (such as health education materials), and participation by the target population. These outputs are expected to lead to the achievement of short-, intermediate-, and long-term outcomes or impacts, seen at the right of the model. (We use outcome and impact interchangeably because health education, communications, public administration, and other relevant fields define these terms differently and sometimes as the opposite of each other.)

This type of logic model is particularly helpful in well-established public health programs that are driven by clear objectives, where the standards of practice and the theory of change are well understood. Immunization and sanitation offer excellent examples. In these programs, it is rarely necessary to revisit the mechanisms of change whereby the goals are to be achieved. At the same time, public health is often presented with new challenges, even in the best-established programs. Under those conditions, the programs may want to

Outputs: the products and activities of the program

revisit a theory of change to address these new challenges.

The second generic model presents the theory of change to a greater extent than the first, employing as its centerpiece if-then statements: If certain inputs and activities occur, then certain outputs and outcomes will occur (22). (These are sometimes called so-that statements.) This generic model specifies the measurement questions to be asked for each element of the logic model and the data sources to be used to answer the questions (22). This type of graphic depiction will be helpful particularly in areas of public health practice that require the active participation of the target population, the active engagement of other organizations, or the need to address new problems facing public health. Under these conditions, the assumptions underlying the theory of change require more discussion and testing.

These generic models are deliberately oversimplified. Simple displays are vital to make sure that users can understand the logic models and respond to them. The whole idea is to provide stakeholders with something they can readily understand and modify in light of their understanding. Logic models are like pie crusts— made to be broken. The evaluability assessment team will know that it is on the right track when stakeholders want to change the logic model. Of course, the team will need to get closure on the logic model at some point.

The two logic models are also oversimplified because it is rare that any single input or set of inputs leads to a single output and a single outcome. Rather, the inputs are complex packages of activities and resources, which generally lead to several outputs and outcomes. The reader will often see such clusters of inputs, outputs, and outcomes in logic models (for example, in **Figure 3** below). Simplifying the display of complex programs is a critically important challenge. For methods to improve the display of complex information, the reader should consult Tufte (45, 46).

Critics object to these generic models on two counts. First, the models are overly linear, whereas many health and social problems are embedded in complex systems. Outcomes may feed back to influence program inputs and outputs, and the impact of the program may reverberate through the system in unanticipated or indirect ways (6, 25). Yet these graphic depictions do not need to reflect every possible consideration.

Second, many program goals are diffuse, and many stakeholders may be left out of the planning process. Therefore, the logic model may not reflect their input. To address these problems, Trochim and his colleagues (44) developed concept mapping, a process to improve stakeholder participation in evaluation planning. Concept mapping recruits a broad range of participants to brainstorm about the important features of the program and potential foci for evaluation. It then structures the relationships among these features, representing these relationships on a map to see how they might cluster together. The map is interpreted for participants who then discuss how these concepts can be used to inform evaluation planning. Trochim and colleagues report that concept mapping has helped to gain closure about evaluation planning in public health contexts that require broad coalitions and have a political dimension. The method appeals to public health professionals because it is inclusive and participatory. However, it can be an elaborate and lengthy process. Respondents can find it burdensome because they are expected to participate in three rounds of nominating and rating program or evaluation foci on two dimensions (importance and feasibility). For many programs it is simply not necessary. The developers caution that it is not appropriate in situations where stakeholders can easily reach a consensus on their own. Also, concept mapping does not replace evaluability assessment because it does not incorporate the same kind of reality testing represented by scouting the program.

Plausibility Analysis

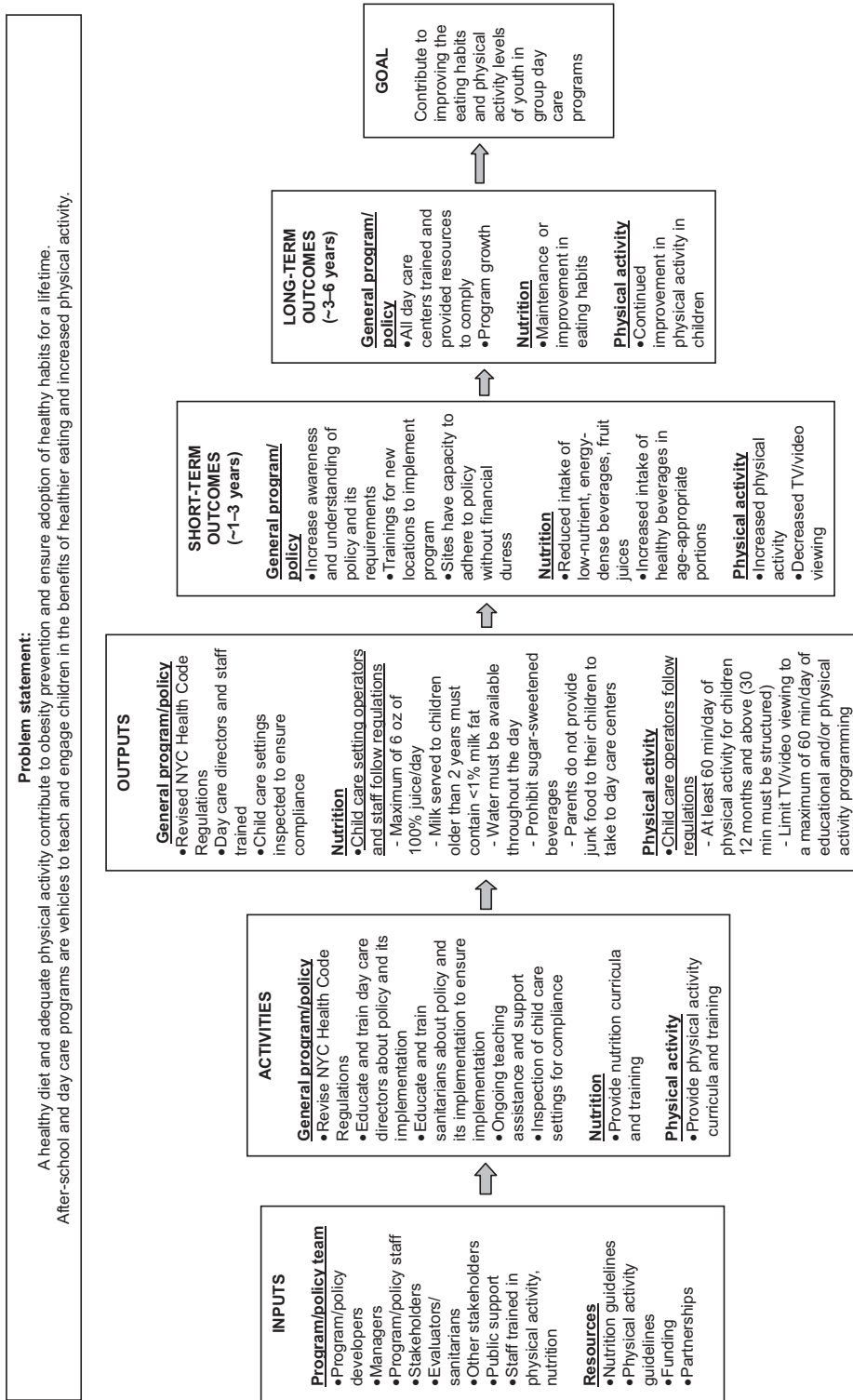Evaluability assessment is distinct from other methods because it addresses whether the

Problem statement:

A healthy diet and adequate physical activity contribute to obesity prevention and ensure adoption of healthy habits for a lifetime.
After-school and day care programs are vehicles to teach and engage children in the benefits of healthier eating and increased physical activity.

INPUTS

Program/policy team
- Program/policy developers
- Managers
- Program/policy staff
- Stakeholders
- Evaluators/ sanitarians
- Other stakeholders
- Public support
- Staff trained in physical activity, nutrition

Resources
- Nutrition guidelines
- Physical activity guidelines
- Funding
- Partnerships

ACTIVITIES

General program/policy
- Revise NYC Health Code Regulations
- Educate and train day care directors about policy and its implementation
- Educate and train sanitarians about policy and its implementation to ensure implementation
- Ongoing teaching assistance and support
- Inspection of child care settings for compliance

Nutrition
- Provide nutrition curricula and training

Physical activity
- Provide physical activity curricula and training

OUTPUTS

General program/policy
- Revised NYC Health Code Regulations
- Day care directors and staff trained
- Child care settings inspected to ensure compliance

Nutrition
- Child care setting operators and staff follow regulations
  - Maximum of 6 oz of 100% juice/day
  - Milk served to children older than 2 years must contain <1% milk fat
  - Water must be available throughout the day
  - Prohibit sugar-sweetened beverages
  - Parents do not provide junk food to their children to take to day care centers

Physical activity
- Child care operators follow regulations
  - At least 60 min/day of physical activity for children 12 months and above (30 min must be structured)
  - Limit TV/video viewing to a maximum of 60 min/day of educational and/or physical activity programming

SHORT-TERM OUTCOMES (~1–3 years)

General program/policy
- Increase awareness and understanding of policy and its requirements
- Trainings for new locations to implement program
- Sites have capacity to adhere to policy without financial duress

Nutrition
- Reduced intake of low-nutrient, energy-dense beverages, fruit juices
- Increased intake of healthy beverages in age-appropriate portions

Physical activity
- Increased physical activity
- Decreased TV/video viewing

LONG-TERM OUTCOMES (~3–6 years)

General program/policy
- All day care centers trained and provided resources to comply
- Program growth

Nutrition
- Maintenance or improvement in eating habits

Physical activity
- Continued improvement in physical activity in children

GOAL

Contribute to improving the eating habits and physical activity levels of youth in group day care programs

Figure 3

New York City's day care logic model. Nutrition, physical activity, and TV-viewing regulations for child care settings in New York City.

Plausibility analysis:
Assessment of the
likelihood that
outcomes are
achievable given the
program's timeframe,
resources, activities,
and context

outcomes of a program are plausible. Plausibility analysis involves assessing whether the outcomes are likely to be achievable given the timeframe and the resources that are available; how the resources are allocated to activities; how the activities are implemented (i.e., with what degree of consistency and fidelity to the original plans, and across program sites if there are multiple sites); and the context within which the program is operating. To assess programs' plausibility to achieve outcomes, one can employ relevant literature, the expertise of program managers and public administrators, reviews of other programs, and common sense. As programs shift over time and as the context shifts around them, it is important to continue to reassess program plausibility and realign the programs as needed, revising goals and outcomes, restructuring activities, or reallocating resources. Analyzing the plausibility of a program to achieve its objectives is critical to evaluability assessment because without it neither the logic model nor analysis of activities has much meaning. It is the core of determining whether something can be evaluated.

Follow-On Steps

Although evaluability assessment is supposed to prepare the way for more formal evaluation, Rog (37) found that most often it did not do so. The act of conducting an evaluability assessment in and of itself may provide enough guidance and program attention in some instances to replace a more thorough evaluation. In fact, the final result may simply be an honest conversation with program stakeholders about the need to revise goals or activities, the flaws of the logic model, or the need to refine program implementation. One way to consider evaluability assessment is like a game of Chutes and Ladders®. Evaluators progress nicely across the game board, but if they land on the squares labeled "stakeholders disagree fundamentally on program goals" or "implausible program," they are back to square 1, or the game may even come to an end. If they land

on the squares labeled "inadequate resources" or "inappropriate activities," they move back to the square labeled "program development." They may also discover that data collection is not feasible or that users cannot agree on a meaningful set of measures. Under any of these circumstances, they may decide not to conduct evaluation.

Stakeholders may not continue with a formal evaluation, even in situations in which the program or elements of the program are considered robust enough to withstand an evaluation. In these situations, it may be that the evaluability assessment provided enough information on the logic of the program and its plausibility in achieving outcomes that decision makers were satisfied to continue their efforts without additional information on outcomes. This is not a desirable use of evaluability assessment because it is not designed for a rigorous test of program outcomes. However, in tight budget situations, knowing that a program is structured and operated in a way that the outcomes are plausible may suffice until additional funds are ready to conduct a rigorous test of program effectiveness.

ILLUSTRATIVE EXAMPLES

Example 1: An Implausible Program that is Not Ready for Evaluation

For a classroom assignment in evaluation, four students in a Master of Public Health degree program conducted an evaluability assessment of a pilot education program directed by a local nonprofit organization (18). This example is noteworthy because it illustrates that (*a*) even novices can conduct excellent evaluability assessments with proper guidance; (*b*) evaluation training can be combined with service learning that benefits local agencies; and (*c*) evaluability assessments can be useful even when programs are not promising, stakeholders disagree about goals, and nothing is ready for evaluation. The anonymous program aimed to educate elementary school children about the function of various organs in their bodies. The assumption

was that such education would lead to healthy lifestyles to keep those organs healthy.

The students collected documentation and interviewed the program manager, the funding organization's leaders, and school personnel who had organized field trips to the program. The students quickly realized that the various stakeholders held very different views of the program's aims. The manager thought it would represent good health education, whereas the teachers believed that it would assist them with the science curriculum. The funder thought it would be good public relations for the non-profit organization. The students brought these disagreements to the stakeholders' attention but found no resolution. Nevertheless, the students developed a logic model that represented those various aims and observed the program in action. They concluded that the activities and exhibits were not plausible to encourage healthy lifestyles, although they might achieve the other goals. Several recommendations aimed to increase the program's potential to improve lifestyles. The students also suggested some data-collection strategies that would be simple and cheap to inform the variety of aims that stakeholders had in mind.

Example 2: A Highly Plausible Policy that is Ready for Evaluation

The CDC in partnership with the Robert Wood Johnson Foundation (RWJF) and ICF Macro conducted evaluability assessment of New York City's day care regulations to prevent childhood obesity (19). This example is noteworthy because (*a*) it clearly meets all the criteria for cost-effective follow-on evaluation; (*b*) it illustrates how evaluability assessment can be applied to policy and environmental interventions for public health (4, 14); and (*c*) it represents a highly promising, real-world public health intervention to translate practice into research (8). The regulation sets standards for physical activity and food offerings in New York City day care environments and sets limits on time spent watching television on site. The regulations are of interest nationally because so few

standards exist for these features of day care settings. The regulations are enforced through inspection of the day care facilities by health department sanitarians and staff of the New York City Bureau of Child Care. The policy affects ~300,000 children in ~1600 group day care centers.

The lead contact for evaluability assessment of the policy was Cathy Nonas, Director of Physical Activity and Nutrition Programs, who had primary responsibility to implement the regulations at the New York City Department of Health and Mental Hygiene. We are grateful for her participation in this work. The CDC contacted her and asked for detailed background information on the activities and goals of the initiative. The evaluation team requested information to identify the key individuals who were closely affiliated with the design and implementation of the initiative, important partners for the initiative, and representatives of the audience touched by the initiative. A single site visit was scheduled because the evaluability assessment team was headquartered in Atlanta. More commonly, evaluability assessments that are performed locally might include several visits to a program site. Before the site visit, the evaluability assessment team used the background materials to develop a preliminary logic model describing the apparent inputs, activities, and desired outcomes of the regulations. The team visited with program staff over two and a half days to learn more about the day care regulations' development, implementation, and desired outcomes. The site visitors spoke with various stakeholders of the policy and observed a couple of the day care center locations where the policy was being enacted. On the last half-day of the site visit, the evaluability assessment team shared preliminary findings with the lead contact and internal stakeholders that she invited to attend. These findings concerned the strengths and weaknesses that the team heard from interviewees about the policy as currently implemented. Also, the site visitors reviewed in detail the logic model to confirm their fuller understanding of the policy. Finally, site visitors used the logic model to

discuss potential opportunities for further data collection and specific evaluation questions that would be useful to address. **Figure 3** reflects the final logic model that emerged from this process. Unlike the generic examples of **Figures 1** and **2**, single resources and activities do not lead to single outputs and outcomes. Rather, each of the outcome variables at each phase is the product of the combined effects of previous stages.

The team prepared a detailed report on the policy's development, current implementation, and evaluation activities under way and possible in the future. The lead contact in New York City reviewed the report and made factual corrections as needed. The site visitors also prepared a set of recommendations for the health department on the underlying theory or logic, the progress of implementation, and evaluation opportunities. A follow-up call with stakeholders allowed a discussion of the findings and preliminary recommendations. The report concluded that (*a*) the policy is fully implemented, (*b*) activities and resources are adequate for implementation, (*c*) the policy is feasible and acceptable to stakeholders, (*d*) it is highly plausible that the policy will contribute to improvements in children's day care environments and therefore contribute to their healthy weight. The report also concluded that the policy was ready for evaluation and outlined some options for conducting such an evaluation (19).

These discussions led to a formal evaluation of the day care regulations, co-funded by the RWJF and the CDC and co-led by investigators at the CDC, New York City Department of Health and Mental Hygiene, and New York University. As of this writing, in 2009, Phase I is under way and concerns the factors affecting day care centers' compliance with the regulations. Phase II will compare effects on children's eating and physical activity while attending day care for centers with high versus low compliance with the regulations. Phase I capitalizes on a health department administrative decision to provide day care centers with technical assistance for good nutrition and age-appropriate physical activity.

The health department established district offices in three areas of the city with a high concentation of poor and nonwhite residents, and these three districts received priority for the technical assistance. However, many other city day care centers serve children with similar low income and demographic makeup and have not yet received the technical assistance. This situation allows a high-quality design to determine whether technical assistance to day care centers enables better compliance with the regulations (19).

BENEFITS OF EVALUABILITY ASSESSMENT FOR PUBLIC HEALTH

Assisting the Planning and Assurance Core Functions

Evaluability assessment can assist the core functions of public health (12), and for a reasonable price. A deeply ingrained public health practice at all levels is the use of goals and objectives for planning. Objectives generally include "how much," "of what," "for whom," and "by when." Each of these components might carry assumptions that need to be tested. Evaluability assessment can reveal whether the assumptions behind these objectives are plausible, and whether stakeholders disagree about them. Where the assumptions are not plausible, a revision is needed, consistent with planning activities. Also ingrained in public health professionals is the fiscal and moral concern to assure the implementation of programs. Evaluability assessment helps to strengthen and possibly assure good implementation by focusing stakeholders' discussions about the best ways to assess progress and performance.

Public health has several planning models that offer systematic steps to build in evaluation. For example, the PRECEDE-PROCEED model uses the diagnostic assessment of needs as a baseline for evaluation (10). The Assessment Protocol for Excellence in Public Health (APEXPH) addresses organization and

management concerns and relations with the community, so it is compatible with the focus on consulting stakeholders, as well as the adequacy of activities and resources (27). The RE-AIM model of behavior change for health promotion and disease prevention also incorporates evaluation into planning (9). Although all these models are useful, they are complementary to evaluability assessment, focusing as it does on the plausibility that resources and activities can achieve objectives and the feasibility of evaluation measurement and design.

Building Evaluation Capacity in Public Health

State and local public health agencies, as well as nonprofit and community-based organizations, are strengthened by having evaluation capacity (3). This capacity improves the programs they run and holds them accountable for federal, state, and local funds they receive. A positive evaluation can also promote the sustainability of programs, including prevention programs, as seen in a recent study funded by the RWJF (17). Of 112 health and health care programs that ended RWJF funding during the period 1986–2001, 80% were still operating in 2002. Among the managers of the surviving programs, 50% believed that a positive evaluation was important to promote program sustainability. Although readers may be disappointed that only 50% gave credit to evaluation, 50% seems high to the staff of many nonprofits and foundations.

In May of 1997, at the instigation of the director, the CDC published the *Framework for Program Evaluation in Public Health* (3). Although it does not name evaluability assessment explicitly, the CDC Framework is built from the same core evaluation principles and practices. The CDC Evaluation Framework is composed of six steps:

1. Engage stakeholders.
2. Describe the program and agree on program goals and expected effects, activities or strategies engaged, resources for implementation, capacity to effect change, stage of development, and contextual information.
3. Focus the evaluation, including the high-priority evaluation questions and designs that maximize confidence in those answers.
4. Gather credible evidence including the data needed to answer evaluation questions, the sources from which those data can be obtained, and the credibility of the data and data sources to the stakeholders.
5. Justify conclusions using standards, analytic procedures, interpretation, and judgment to determine the extent to which evaluation conclusions are justified.
6. Ensure use and share lessons learned.

The first four steps are directly relevant to evaluability assessment. One significant difference between evaluability assessment and the CDC Framework is that evaluability assessment explicitly asks the practitioner to question what is learned from the steps and to ask which evaluation activities are appropriate, given that information. The Framework amply describes what evaluators need to consider at each of these steps, but it is silent on how to think about them or the implications of what is learned once they have been considered. For example, when focusing the evaluation design, the CDC Framework highlights the need to identify the data to measure the constructs of interest, the sources of those data, and a design that maximizes our confidence in the results. No explicit guidance appears about assessing the feasibility of designs or the accessibility and affordability of data collection. In contrast, evaluability assessment requires asking these questions. In fact, if the preferred designs and data-collection methods are not feasible, some additional evaluation planning may be required, perhaps even rethinking the evaluation questions. The CDC Evaluation Framework and evaluability assessment are highly complementary approaches that, used together, can increase the number of evaluations that are appropriately tailored and likely to yield valuable and usable information.

Navigating Federal Performance Measurement Requirements

Evaluability assessment is consistent with best practices in public administration. For this reason, it can help meet the reporting requirements of GPRA and PART (30, 31). President Obama highlighted performance measurement in his inaugural address, making it likely that federal requirements will continue and may become even more influential. Wholey and his colleagues played a major role in implementing GPRA and in the movement to increase federal performance measurement (33, 49). Clearly there are similarities among the approaches used in evaluability assessment and GPRA, and the specific methods that PART requires.

GPRA requires that federal agencies develop five-year plans that include a mission statement and long-term goals. The agencies are to provide an annual performance plan setting forth goals for the year, how the agencies are going to meet those goals, and how the goals can be measured. An annual performance report reviews the extent to which the agency met the performance goal. PART expands the scope of the GPRA requirements by evaluating program progress toward long-term, ambitious health outcome goals relative to current funding levels. PART also assesses program accountability by reviewing management and operations practices, including workforce planning, management of information technology, and financial management systems. Like evaluability assessment, PART proposes to conduct a high-level preliminary assessment to (a) measure and diagnose program performance; (b) evaluate programs in a systematic, consistent, and clear manner; (c) help the agency and OMB make informed decisions for management, budget, and legislative or regulatory improvements; (d) focus on program improvements; and (e) measure program progress over time.

Both juvenile justice and mental health programs have used evaluability assessment to prepare for GPRA and PART reviews (13). The Community Mental Health Services Block Grant Program offers one example of an evaluability assessment for this purpose (26). No evaluations had been conducted of the block grant since its inception in 1981. However, in 2001, a PART Corrective Action Plan requested an evaluation. Given the program's evolution and complexity, an evaluability assessment was commissioned and awarded as a first step to plan the evaluation. The assessment revealed that local and federal stakeholders did not agree on the goals of the program, the relationships between federal and state activities were not clear, and both the activities and the resources were too limited to have any plausible effects on mental health outcomes at a population level. However, the evaluability assessment facilitated discussions among stakeholders and gained their buy-in to develop appropriate performance measures. The final logic model portrayed the limited scope of the block grant, reduced expectations about impact, focused on the shared responsibility of federal and state levels for outcomes, and facilitated the creation of an evaluation design that all could accept. (See sidebar, NIOSH Responds to the PART Requirements.)

Translating Research into Practice

Some new applications of evaluability assessment are useful for translating research into practice. In this case, unlike the examples we offered above, there exist good studies of the effectiveness of interventions, which are then presumed to be ready for dissemination and uptake by public health, medical, and community organizations. However, such evidence-based practices need to be realistic and feasible for practitioners to incorporate them into the management of programs (8). As practitioners implement an evidence-based intervention or program, evaluability assessment can test assumptions about the program's operation in new settings for a reasonably low cost and in real time. Where adjustments are indicated, the method can guide adaptation to real-world considerations. Where the practice cannot be implemented because of constraints

on the setting (inputs) or acceptability to the target populations, evaluability assessment will contribute an early warning for a low cost.

One illustration, and a process that borrowed from evaluability assessment, is the organizational capacity assessment that was undertaken for the CDC-funded Community Intervention Trial for Youth, a 13-city community-randomized trial designed to evaluate a multicomponent, community-based HIV prevention program for young men who have sex with men (23). The program components were evidence based but derived from studies in more controlled, better-resourced settings with mostly older men. The randomized experiment was necessary to establish effectiveness across cities; the organizational assessment, however, established in advance which community-based organizations could achieve the necessary program outputs (prescribed activities, participation of the young men).

Thus evaluability assessment can extend our understanding of external validity—the populations, organizations, settings, and treatment adaptations to which evidence of effectiveness can and cannot be generalized (16). If the evaluability assessment indicates that the intervention is not acceptable to the population, then it is unlikely to gain the necessary participation; if personnel must adapt the intervention, then it is imperative to discuss whether the adapted outputs are still consistent with the evidence-based model. Note that evaluability assessment can be fairly definitive in ruling out the settings and populations where an evidence-based practice is not plausibly effective ("You can't get there from here"). The challenge comes with ruling in whether a practice will be effective in settings, organizations, populations, and treatment adaptations that are substantially different from those in the original studies. External validity relies on an inductive process, unlike the deductive process of assessing internal validity: We can rely on theory, context, and common sense, but without an independent test of effectiveness in the new context, we cannot be definitive. Evalua-

NIOSH RESPONDS TO THE PART REQUIREMENTS

The National Institute for Occupational Safety and Health (NIOSH) conducts research to improve occupational health and safety in the United States. In 2004, NIOSH requested a committee of the Institute of Medicine to evaluate NIOSH research programs for their relevance and impact (6). Although NIOSH needed these evaluations to satisfy the OMB-mandated PART process, NIOSH leaders chose program improvement as their primary aim. By 2009, eight expert panels had evaluated programs such as hearing loss prevention and mining safety and health. The Institute's committee utilized techniques related to evaluability assessment, which is not surprising because Wholey was a committee member. The committee constructed and revised an evaluation framework and a logic model. The expert panels consulted stakeholders and assessed whether program activities and resources matched the research programs' objectives. Over and above these requirements, the expert panels assessed program resources, activities, and products in light of the nation's priorities for occupational health and safety. The evaluations identified management practices that needed adjustment. NIOSH leadership intends to use the framework to continue improving NIOSH research programs.

bility assessment can reduce uncertainty about effectiveness in new contexts and help us make a reasonable conjecture about whether the evidence-based practice will generalize to those contexts.

Translating Practice into Research: The Systematic Screening and Assessment Method

A new development for evaluability assessment is to translate practice into research (8) by identifying promising practices that are ready for evaluation. The process, which we have termed the systematic screening and assessment (SSA) method, aims to cast a wide net for promising innovations, then screen these innovations systematically to assure that those remaining at each step have a high likelihood of being effective and ready for evaluation. The process involves (*a*) soliciting a topic or theme, such as

access to healthy food in poor neighborhoods; (b) soliciting a high volume of innovations that address this theme through a broad-based scan; (c) using an expert panel of researchers and content experts to screen these nominations for those that are most promising (i.e., plausible to have large effects, to reach large numbers of the target group, to generalize to other populations and settings, to be feasible, acceptable, and sustainable); (d) conducting evaluability assessments of those that pass the expert panel screen; (e) engaging an expert panel to review the evaluability assessment reports; and (f) identifying those innovations that are both promising and ready for evaluation.

The SSA method is particularly important in areas such as childhood obesity prevention, where few effective interventions are available and the mechanisms underlying change are not well understood (4, 14). Also, a recent focus in prevention is on policy and environmental interventions. Whereas a great deal is known about policy and environmental interventions for tobacco control, relatively little is known about this strategy for childhood obesity prevention (4, 14).

A recent RWJF/CDC initiative used the SSA method to identify promising practices in childhood obesity (19). Of the 458 innovations nominated in this initiative, 174 met inclusion criteria: 53 of these were sufficiently promising to merit evaluability assessments, and 20 were deemed to be the most promising and ready for evaluation. An example of the yield from this process is the evaluability assessment of the New York City day care regulations, described above (19). Other examples include the Food Trust's Fresh Food Financing Initiative, which has brought more than 30 new supermarkets to low-income areas of Philadelphia; an incentive program for WIC recipients to buy fresh fruit and vegetables; and a national effort to build playgrounds in city neighborhoods that lack such facilities (19). Five of these innovations are now funded for evaluation.

The SSA method differs from stand-alone single-site evaluability assessments because it assumes from the outset that most innovations are unlikely to pass the early screening process (plausible for *large* effects reaching *large* numbers); therefore, a high volume of nominations is needed. In a typical evaluability assessment—and most evaluations for that matter—the program to be studied is selected by a client or a group of stakeholders with limited prior information about effectiveness (18, 29, 35, 37, 40–43, 47, 48). The SSA method structures the process to maximize prior information about promise and readiness for evaluation and to select innovations on that basis. This is a cost-effective strategy for public health. The total cost was $2.4 million, and the cost per promising innovation was $120,000, a relatively inexpensive method considering the substantial cost of no-effect conclusions in evaluation (7, 20, 35, 41, 47).

The SSA method provided other benefits. The 39 nominations and 7 evaluability assessments of comprehensive school-based physical activity programs identified some problems typical of implementation, as well as some good practices that would promote increased physical activity in schools. This information led directly to a CDC guide, under production in 2009, to improve such programs. Another unexpected outcome of using the SSA method in childhood obesity was that the expert panelists shifted or reframed their assumptions, expectations, and knowledge of childhood obesity prevention. For example, they initially assumed that school wellness policies as mandated by the Congress would have an effect on the foods that schools provided to children and on children's opportunities to be physically active. However, of 146 policies that were nominated, only 4 were rated as promising to achieve these goals. Together with other information, this process made the panel much less optimistic about the power of these policies (19). In contrast, the panel was initially skeptical about the power of farmers markets to change access to healthy foods. As they learned more about the actual practices involved, they revised their opinion about the promise of this strategy. Also, evaluability assessments stimulated the expert panel members to discuss the potential underlying mechanisms

for change, the viable designs and measures, and appropriate expectations of programs and policies. In this way, the SSA method built on and strengthened the research program to address childhood obesity (19).

Finally, the SSA method was explicit in providing constructive feedback to each of the 48 innovations that underwent evaluability assessment. A survey of the innovation developers and managers indicated they generally found the process useful and often intended to use the feedback for program improvement (19). We have often observed that program development is a result of the feedback process in evaluability assessment, yet this aspect goes largely unrecognized in the literature. A previous CDC initiative to discover promising obesity-prevention efforts in work sites determined that evaluability assessment was an excellent vehicle for program development (11). In both obesity prevention initiatives, evaluability assessments gave something back to the innovations that were studied by offering technical assistance from CDC experts.

CONCLUSION

Evaluability assessments are a cost-effective strategy to assure that limited evaluation resources can be used in the most appropriate ways. The method can benefit research and practice in public health through program development, evaluation capacity building, performance measurement, assessment of the external validity of evidence-based practices, and identification of promising practices that merit more formal evaluation. Although the process is supposed to result in evaluation, it often does not do so, but for good reason: Program development has to come first.

DISCLOSURE STATEMENT

L.L. is employed by the Robert Wood Johnson Foundation and is a member and former president of the American Evaluation Association.

LITERATURE CITED

1. Am. Eval. Assoc. 2004 [1994]. *Guiding principles for evaluators*. **http://www.eval.org/Publications/ GuidingPrinciples.asp**
2. Carman JG, Fredericks KA. 2008. Nonprofits and evaluation: empirical evidence from the field. In *Nonprofits and Evaluation*, ed. JG Carman, KA Fredericks. *New Dir. Eval.* 2008(119):51–72
3. Cent. Dis. Control Prev. 1999. Framework for program evaluation in public health. *MMWR* 48 (No. RR-11):1–40
4. Cent. Dis. Control Prev. 2009. *The Guide to Community Preventive Services*. **http://www. thecommunityguide.org/index.html**
5. Collins CB, Sloop K, Napp D. 2004. *A strategy to enhance evaluation capacity for HIV prevention providers*. Presented at Int. Conf. AIDS, 15th, Bangkok, Thailand (Abstr. No. TuPeE5493)
6. Comm. Rev. NIOSH Res. Progr. 2009. *Evaluating Occupational Health and Safety Programs: Framework and Next Steps*. Washington, DC: Natl. Acad. Press
7. Cronbach LJ, with Shapiro K. 1982. *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass
8. Glasgow RE, Green LW, Klesges LW, Abrams DB, Fisher EB, et al. 2006. External validity: We need to do more. *Ann. Behav. Med.* 31:105–8
9. Glasgow RE, Linnan LA. 2008. Evaluation of theory-based interventions. In *Health Behavior and Health Education: Theory, Research and Practice*, ed. K Glanz, BK Rimer, K Viswanath, pp. 487–508. San Francisco: Jossey-Bass
10. Green LW, Kreuter M. 2004. *Health Program Planning: An Educational and Ecological Approach*. New York: McGraw-Hill

11. Hersey J, Williams-Peihota P, Sparling PB, Alexander J, Hill MD, et al. 2008. Promising practices in promotion of healthy weight at small and medium-sized US worksites. *Prev. Chron. Dis.* 5(4). **http://www.cdc.gov/PCD/issues/2008/oct/pdf/07_0172.pdf**

12. Inst. Med. 1988. *The Future of Public Health*. Washington, DC: Natl. Acad.

13. Justice Res. Stat. Assoc. 2004. *Approaches to assessing juvenile justice program performance*. **http://www.jrsainfo.org/pubs/juv-justice/approaches_assessing.pdf**

14. Koplan JP, Liverman CT, Kraak VA. 2005. *Preventing Childhood Obesity: Health in the Balance*. Washington, DC: Natl. Acad.

15. Leviton LC. 2001. Building evaluation's collective capacity: American Evaluation Association Presidential address. *Am. J. Eval.* 22:1–12

16. Leviton LC. 2001. External validity. In *The International Encyclopedia of the Social and Behavioral Sciences*, ed. NJ Smelser, PB Baltes, 8:5195–200. Oxford: Elsevier

17. Leviton LC, Bass M. 2004. Using evaluation to advance a foundation's mission. In *Foundations and Evaluation: Contexts and Practices for Effective Philanthropy*, ed. M Braverman, N Constantine, JK Slater, pp. 3–26. San Francisco: Jossey-Bass

18. Leviton LC, Collins C, Laird B, Kratt P. 1998. Teaching evaluation using evaluability assessment. *Evaluation* 4:389–409

19. Leviton LC, Kettel Khan L, Dawkins N. 2010. *The Systematic Screening and Assessment Method: Identifying Practices Worth Evaluating*. New Dir. Eval. 2010(125)

20. Lipsey MW. 1988. Practice and malpractice in evaluation research. *Eval. Pract.* 9:5–25

21. Lipsey MW. 1993. Theory as method: small theories of treatments. In *Understanding Causes and Generalizing About Them*, ed. LB Sechrest, AG Scott. *New Dir. Progr. Eval.* 1993(57):1–82

22. McLaughlin JA, Jordan GB. 2004. Using logic models. See Ref. 49, pp. 7–32

23. Miller RL, Bedney BJ, Guenther-Grey C. 2003. Assessing organizational capacity to deliver HIV prevention services collaboratively: tales from the field. *Health Educ. Behav.* 30:582–600

24. Moran G. 2005. *Assessing the effectiveness of discharge planning to prevent homelessness: an evaluability assessment*. Presented at Joint Meet. Can. Eval. Soc. Am. Eval. Assoc., Toronto

25. Morell JA. 2005. Why are there unintended consequences of program action, and what are the implications for doing evaluation? *Am. J. Eval.* 26:444–63

26. Mulkern V. 2005. *Evaluability assessment of the Community Mental Health Block Grant Program*. Presented at Joint Meet. Can. Eval. Soc. Am. Eval. Assoc., Toronto

27. Natl. Assoc. County City Health Off. 1991. *The Assessment Protocol for Excellence in Public Health (APEXPH)*. Washington, DC: Natl. Assoc. County City Health Off.

28. Natl. Prev. Inf. Netw. 2009. *HIV Prevention Program Evaluation—Materials Database*. **http://www.cdcnpin.org/scripts/dhap/list.asp**

29. Nay JN, Kay P. 1982. *Government Oversight and Evaluability Assessment*. Lexington, MA: Lexington

30. Off. Manag. Budget. 2009. *Assessing program performance*. **http://www.whitehouse.gov/omb/part/**

31. Off. Manag. Budget. 2009. *Government Performance Results Act of 1993*. **http://www.whitehouse.gov/omb/mgmt-gpra_gplaw2m/**

32. OMG Cent. Collab. Learn. 2007. *Brief Assessments of Community-Based Childhood Obesity Prevention Within the Injury Free Coalition for Kids Initiative Sites: Final report*. Philadelphia: OMG Cent. Collab. Learn. **http://www.rwjf.org/files/research/37810ifckevalfinal.pdf**

33. Poister TH. 2004. Performance monitoring. See Ref. 49, pp. 98–125

34. Preskill H, Jones N. 2009. *A Practical Guide for Engaging Stakeholders in Developing Evaluation Questions*. Boston: FSG Social Impact Advisors. **http://www.fsg-impact.org/advice/Strategic_Evaluation/index.htm**

35. Reichert CS. 1994. Summative evaluation, formative evaluation, and tactical research. *Eval. Pract.* 15:275–81

36. Rhodes S. 2007. *Experiences with Exploratory Evaluation: A Summary of Lessons Learned*. Winston-Salem: Wake Forest Univ. Health Sci.

37. Rog DJ. 1985. *A methodological analysis of evaluability assessment*. PhD thesis. Vanderbilt Univ. 166 pp.

38. Rog DJ. 2005. *Evaluability assessment: then and now*. Presented at Joint Meet. Can. Eval. Soc. Am. Eval. Assoc., Toronto

39. Rog DJ, Gutman M. 2007. *Brief Assessments of Environmental and Policy Interventions on Healthy Eating: Lessons Emerging about the Methodology*. Princeton, NJ: Gutman

40. Rutman L. 1980. *Planning Useful Evaluations: Evaluability Assessment*. London: Sage

41. Shadish WR, Cook TD, Leviton LC. 1991. *Foundations of Program Evaluation: Theorists and Their Theories*. Newbury Park, CA: Sage

42. Smith MF. 1989. *Evaluability Assessment: A Practical Approach*. Boston: Kluwer Acad.

43. Trevisan MS. 2007. Evaluability assessment from 1986 to 2006. *Am. J. Eval.* 28:209–303

44. Trochim W, Milstein B, Wood B, Jackson S, Pressler V. 2004. Setting objectives for community and systems change: an application of concept mapping for planning a statewide health improvement initiative. *Health Promot. Pract.* 5:8–19

45. Tufte ER. 1997. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press. 2nd ed.

46. Tufte ER. 1997. *Visual and Statistical Thinking: Displays of Evidence for Decision Making*. Cheshire, CT: Graphics Press

47. Wholey JS. 1979. *Evaluation: Promise and Performance*. Washington, DC: Urban Inst.

48. Wholey JS. 2004. Assessing the feasibility and likely usefulness of evaluation. See Ref. 49, pp. 33–62

49. Wholey JS, Hatry HP, Newcomer KE, eds. 2004. *Handbook of Practical Program Evaluation*. San Francisco: Jossey-Bass. 2nd ed.

50. Wilson DB, Lipsey MW. 2001. The role of method in treatment effectiveness research: evidence from meta-analysis. *Psychol. Methods* 6:413–29

Contents

Environmental and Occupational Health

Public Health Practice

Social Environment and Behavior

Health Services

Indexes

Errata

An online log of corrections to *Annual Review of Public Health* articles may be found
at http://publhealth.annualreviews.org/